Reliability Survey of Military Acquisition Systems

Jonathan L. Bell, Ph.D., Institute for Defense Analyses
Matthew R. Avery, Ph.D., Institute for Defense Analyses
Michael C. Wells, Institute for Defense Analyses

Key Words: Survey, Department of Defense, Military, Reliability

SUMMARY & CONCLUSIONS

Improving the reliability of military systems within the Department of Defense (DoD) is a key priority. Test results from the last few decades indicate that the DoD has not yet realized significant statistical improvements in the reliability of many systems. However, there is evidence that those systems that implemented a comprehensive reliability growth program are more likely to meet their development goals. Reliable systems cost less overall, are more likely to be available when called upon, and enable a longer system lifespan. Reliability is more effectively and efficiently designed-in early (design for reliability) vice being tested-in late. While more upfront effort is required to build reliable systems, the future savings potential is too great to ignore.

At the request of the Director, Operational Test and Evaluation (DOT&E), the Institute for Defense Analyses (IDA) has conducted annual reliability surveys of DoD programs under DOT&E oversight since 2009 to provide a continuing understanding of the extent to which military programs are implementing reliability-focused DoD policy guidance and assess whether the implementation of this guidance is leading to improved reliability. This paper provides an assessment of the survey results.

Overall survey results support the understanding that systems with a comprehensive reliability growth program are more likely to meet reliability goals in testing. In particular, the results show the importance of establishing and meeting Reliability, Availability, and Maintainability (RAM) entrance criteria before proceeding to operational testing (OT). While many programs did not establish or meet RAM entrance criteria, those that did were far more likely to demonstrate reliability at or above the required value during OT. Examples of effective RAM entrance criteria include (1) demonstrating in the last developmental test event prior to the OT a reliability point estimate that is consistent with the reliability growth curve, and (2) for automated information systems and software-intensive sensor and weapons systems, ensuring that there are no open Category 1 or 2 deficiency reports prior to OT. There is also evidence that having intermediate goals linked to the reliability growth curve improves the chance of meeting RAM entrance criteria.

The survey results also indicate that programs are increasingly incorporating reliability-focused policy guidance, but despite these policy implementation improvements, many programs still fail to reach reliability goals. In other words, the policies have not yet proven effective at improving reliability trends. The reasons programs fail to reach reliability goals include inadequate requirements, unrealistic assumptions, lack of a design for reliability effort, and failure to employ a comprehensive reliability growth process. Although the DoD is in a period of new policy that emphasizes good reliability growth principles, without a consistent implementation of those principles, the reliability trend will likely remain flat.

In the future, programs need to do a better job incorporating a robust design and reliability growth program from the beginning that includes the design for reliability tenets described in the ANSI/GEIA-STD-0009, "Reliability Program Standard for Systems Design, Development, and Manufacturing." Programs that follow this practice are more likely to be reliable. There should be a greater emphasis on ensuring that reliability requirements are achievable, and reliability expectations during each phase of development are supported by realistic assumptions that are linked with systems engineering activities. Programs should also establish RAM entrance criteria and ensure these criteria are met prior to proceeding to the next test phase. A program's reliability growth curves should be constructed with a series of intermediate goals, with time allowed in the program schedule for test-fix-test activities to support achieving those goals. Finally, when sufficient evidence exists to determine that a program's demonstrated reliability is significantly below the growth curve, that program should develop a path forward to address shortfalls and brief their corrective action plan to the acquisition executive.

1 INTRODUCTION

DOT&E is the principal staff assistant and senior advisor to the Secretary of Defense on operational test and evaluation (OT&E) in the DoD. DOT&E oversees major DoD acquisition programs to ensure OT&E is adequate to confirm operational effectiveness and suitability of the defense system in combat use [1]. Data from DOT&E reports to congress suggest that despite establishment over the years of policies intended to encourage development of more reliable systems, DoD system reliability has not improved. From 1997 to 2013, only 56 percent of the systems that underwent an OT met or exceeded

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE **2014**	2. REPORT TYPE	3. DATES COVERED **00-00-2014 to 00-00-2014**
4. TITLE AND SUBTITLE **Reliability Survey of Military Acquisition Systems**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Institute for Defense Analyses,4850 Mark Center Drive,Alexandria,VA,22311-1882**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**5**	

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std Z39-18

their reliability threshold requirements [2]. Further analysis suggests there has been no improvement in the fraction of programs meeting their reliability requirements over time.

To better understand these trends, DOT&E requested IDA to conduct a survey of military programs in each of the past five years to determine the extent to which reliability-focused policy guidance is being implemented and to assess whether it is leading to improved reliability. IDA developed a survey and distributed it to research staff members that are subject matter experts on the programs of interest. Survey topics included questions on the program's reliability growth plan, plans for tracking reliability during development, whether the program has a process of calculating the reliability growth potential, and questions on reliability performance in OT. Select survey questions are listed in Table 1. For most questions, respondents were required to answer “yes,” “no”, or “unknown.” Respondents were also provided with opportunities to enter comments for each question.

#	Survey Question
1	What is the program title? (select from a list)
2	What is the lead Service or military department?
3	What acquisition phase is the program in?
4	Has a TEMP been approved for the program in Fiscal Year (FY) 2012?
5	Does the program have a reliability growth or improvement strategy?
5b	Does the test plan describe the reliability growth or improvement strategy or reference where the strategy can be found?
5c	Does the program have reliability growth curves?
5c1	Do the reliability growth curves appear in the TEMP?
5c2	Was the reliability growth curve used to develop intermediate reliability goal(s)?
5c3	Are the reliability growth goal(s) linked to OTs (e.g., IOT&E, FOT&E, and/or MS C Operational Assessments)? In other words, are the reliability goal(s) based on demonstration of the reliability threshold(s) during an OT with statistical confidence (1- consumer risk) and power (1-producer risk)?
6	Does the program routinely perform assessments using reliability metrics to ensure reliability growth is on track to achieve requirements (e.g., assessment conferences to assess fix effectiveness of corrective actions, reliability tracking models to determination if the reliability is increasing with time)?
7	Does the program have a process for calculating the reliability growth potential?
8	Did your program have an operational test in FY12?
8a	What type of operational test was it? (DT/OT, OA/LUT, IOT&E, FOT&E)
8b	Did the program establish and meet RAM based entrance criteria in Developmental Testing (DT)?
8c	Were RAM based exit criteria met?
8d	Did the system demonstrate a reliability at or above the required value during the OT?

Table 1 –Select Survey Questions

The most recent survey was conducted in 2013 and focused on programs that submitted a Test and Evaluation Master Plan (TEMP) to DOT&E or had an OT in FY 2012. The TEMP is the overarching document that describes the program's test plan [3].

1.1 *Survey Analysis Approach*

Analysis of each survey question considered how the responses varied by time by comparing responses in the most recent survey to the earlier surveys by TEMP date. Duplicate survey entries between surveys were removed. The analysis also considered differences by lead service including the Army, Navy, and Air Force (Marine Corps responses were grouped with the Navy), and by acquisition phase.

The analysis binned the responses using the following TEMP date categories to maintain consistency with the methodology used in previous survey analyses:

- Dated before July 2008, prior to approval of a key DoD reliability policy (75 responses)
- Dated between June 2008 and October 2010 (81 responses)
- Dated in FY 2011 (57 responses)
- Dated in FY 2012 or FY2013 13 (52 responses).

Where appropriate, contingency tables were used to record and analyze the relationship between two or more categorical variables. This allowed the determination of whether the observed results were statistically significant.

1.2 *Population of Survey Responses*

IDA analysts completed 97 responses in the most recent reliability survey conducted in 2013. Of the 97 responses, 52 were for programs that had an FY 2012 or 2013 TEMP, 66 were for programs that had an FY 2012 OT, and 7 were for programs that did not have an FY 2012 or 2013 TEMP or OT. Of the 66 programs with an FY 2012 OT, 28 also had an FY 2012 or 2013 TEMP. Table 2 shows the breakdown of responses by acquisition phase, lead Service, and test type. Approximately 63 percent of systems represented by survey responses were past their Initial Operational Test (IOT).

<table>
<tr><th colspan="2">Acquisition Phase</th><th colspan="2">Lead Service</th><th colspan="2">Test Type</th></tr>
<tr><td>MSA</td><td>1 (1%)</td><td>Army</td><td>17 (27%)</td><td rowspan="2">DT/OT</td><td rowspan="2">5 (8%)</td></tr>
<tr><td>TD</td><td>2 (2%)</td><td>Navy</td><td>45 (44%)</td></tr>
<tr><td>EMD</td><td>17 (18%)</td><td rowspan="2">Air Force</td><td rowspan="2">26 (27%)</td><td rowspan="2">OA or LUT</td><td rowspan="2">14 (21%)</td></tr>
<tr><td>Pre-IOT&E P&D</td><td>13 (14%)</td></tr>
<tr><td>Post-IOT&E P&D</td><td>43 (47%)</td><td rowspan="2">Marine Corps</td><td rowspan="2">3 (3%)</td><td rowspan="2">IOT&E</td><td rowspan="2">27 (41%)</td></tr>
<tr><td>O&S</td><td>13 (16%)</td></tr>
<tr><td>Other</td><td>2 (2%)</td><td>Other</td><td>6 (6%)</td><td>FOT&E</td><td>20 (30%)</td></tr>
</table>

Acronyms: Materiel Solution Analysis (MSA); Technology Development (TD); Engineering and Manufacturing Development (EMD); Production and Deployment (P&D); Operations and Support (O&S), Limited User Test (LUT), Operational Assessment (OA), Follow-on Operational Test and Evaluation (FOT&E).

Table 2 – Breakdown of Survey Responses by Number of Responses and Percent

2 SURVEY RESULTS

Overall results, based on analysis of survey responses and user comments, reinforce the understanding that systems with a robust reliability growth program are more likely to reach reliability goals. In particular, analysis results revealed the importance of establishing RAM entrance criteria and intermediate goals that are linked to the reliability growth curve. As shown in Table 3, programs that establish and meet their RAM entrance criteria are more likely to demonstrate reliability at or above the required value during OT. Examples of effective RAM entrance criteria include (1) demonstrating, in the last DT event before the IOT&E, a reliability point estimate that is consistent with the reliability growth curve, and (2) for automated information systems, ensuring that there are no open category 1 or 2 deficiency reports prior to OT [4].

		Demonstrated a reliability at or above the required value during IOT&E/FOT&E	Pearson p-value
Met RAM entrance criteria	Yes	87% (13 of 15)	0.0001*
	No	0% (0 of 7)	

Table 3 – RAM Entrance Criteria and Meeting Reliability Thresholds in OT Considering 2013 Survey Responses

Of the 15 programs in Table 3 that established and met their RAM entrance criteria in DT, 13 met their reliability goals in OT. None of the seven programs that failed to meet their entrance criteria in DT went on to meet their reliability thresholds in OT. The Pearson p-value in shown in Table 3 indicates that this result is statistically significant. This result suggests that programs that do well in DT are more likely to so well in later OT. However, despite this obvious result, many programs do not establish RAM entrance criteria, and programs that fail to meet entrance criteria in DT are still permitted to move forward and participate in OT. This result confirms that moving programs forward that perform poorly in DT increases the risk they will fail to reach reliability thresholds in OT.

There is also evidence that programs that have intermediate goals that are linked to the reliability growth curve are more likely to meet their RAM entrance criteria as shown in Table 4.

		Demonstrated a reliability at or above the required value during IOT&E/FOT&E	Pearson p-value
Has intermediate goals linked to the growth curve	Yes	82% (14 of 17)	0.0665
	No	14% (1 of 7)	

Table 4 – Intermediate Goals and RAM Entrance Considering Combined Survey Responses

Overall results also suggest that implementing RAM policies alone, without the support of a robust reliability growth program, is insufficient to improve the chance of success in OT. Analysis of responses collected in 2013 for programs that had an IOT&E or FOT&E provide no significant evidence that implementation of RAM policies alone improves the chance of demonstrating RAM threshold during OT. As shown in Table 5, there was no single policy area that could be correlated with success in OT. In fact, a smaller fraction of programs with growth curves met their RAM entrance and exit criteria compared to programs that do not have reliability growth curves. User comments report a variety of reliability growth plan inadequacies such as requirement deficiencies, policy implementation concerns, and testing limitations.

		Demonstrated a reliability at or above the required value during IOT&E/FOT&E	Pearson p-value
Having a reliability growth (RG) or improvement strategy	Yes	61% (23 of 38)	0.6830
	No	50% (2 of 4)	
Having RG curves	Yes	58% (14 of 24)	0.7173
	No	64% (9 of 14)	
Having intermediate goals linked to the growth curve	Yes	55% (6 of 11)	0.8548
	No	58% (7 of 12)	
Having RG linked to OTs	Yes	57% (8 of 14)	0.8887
	No	60% (6 of 10)	
Uses reliability metrics to ensure RG is on track	Yes	60% (18 of 30)	1.000
	No	60% (3 of 5)	
Calculates the RG potential	Yes	40% (6 of 15)	0.2103
	No	63% (10 of 16)	

Table 5 – Influence of Reliability Policies on Meeting Thresholds in OT Considering 2013 Survey Responses

For example, some respondents commented that reliability growth curves were constructed as an afterthought, retrofitted in the TEMP only after DOT&E requested information on it. In these instances, the construction of reliability growth curve was to comply with a paper policy, rather than to reflect systems engineering activities. Other respondents indicated that the reliability requirements were not achievable, because they were based on faulty modeling assumptions or they were unrealistically high compared to similar system. Finally, some respondents commented that there was insufficient testing in OT to evaluate the reliability requirement or the reliability growth model inputs were not based on realistic assumptions.

Consistent with the result of previous surveys, survey responses collected in 2013 provide no evidence of improvement in the percentage of programs that met their

RAM entrance or exit criteria. Compared to other types of OT, FOT&Es had the highest fraction of programs that met their exit criteria or demonstrated reliability above the requirement (Figure 1). This suggests that many programs do not reach their reliability goals until after fielding.

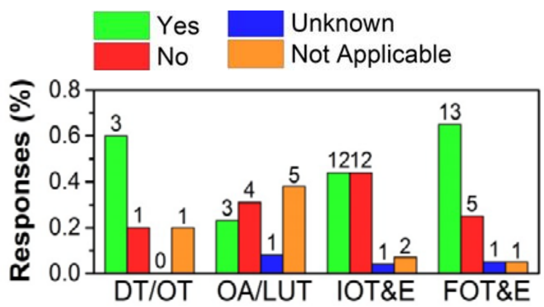

Figure 1 – Fraction/Number of Responses Indicating Whether the System Demonstrated a Reliability at or Above the Required Value During OT by Test Type

2.1 *Comparison of Responses by TEMP Date*

Analysis of responses shows that the fraction of programs that implement reliability-focused policy guidance continues to improve. Areas of continuous policy implementation improvement over time included the following:

- Having a reliability growth (RG) strategy
- Documenting reliability RG in the TEMP
- Incorporating RG curves into the TEMP
- Having a process for calculating RG potential.

The results for these questions are listed in Table 6 for known "Yes" or "No" responses. Analysis results suggest that the improvement over time is statistically significant at the 90 percent confidence level.

	TEMP Approval Date				
Reliability Survey Question	before 07/2008	07/2008-09/2010	2011	2012 or 2013	p-value
Have a RG strategy	55% (35/64)	66% (47/71)	73% (40/55)	92% (48/52)	0.0002*
Document RG strategy in the TEMP	43% (15/35)	77% (36/47)	80% (28/35)	90% (43/48)	<0.0001*
Incorporate RG curves into the TEMP	30% (6/20)	57% (16/28)	68% (15/22)	81% (25/31)	0.0032*
Have a process for calculating the RG potential	17% (13/75)	23% (19/81)	30% (17/57)	40%(21/52)	0.0010*

Table 6 – Improvements in Reliability Policy Implementation Over Time

As shown in Table 7, the fraction of FY 2012 or 2013 TEMP programs that use the reliability growth curve to develop intermediate goals improved (59 percent) compared to FY 2011 TEMP programs (48 percent), but remained below the fraction observed for programs with TEMPs approved between June 2008 and October 2010 (73 percent). The fraction of FY 2012 or 2013 TEMP programs that use reliability metrics to ensure growth is on track to achieve requirements also increased, reaching a higher percentage than that observed for older TEMP date categories.

	TEMP Approval Date				
Reliability Survey Question	before 07/2008	07/2008-09/2010	2011	2012 or 2013	p-value
Use RG curve to develop intermediate goas	44% (8/18)	70% (19/27)	48% (11/23)	59% (19/32)	0.2625
Use reliability metrics to ensure growth is on track to achieve requirements	69% (38/55)	79% (45/57)	64% (27/42)	87% (40/46)	0.0556*

Table 7 – Recent Reliability Improvement Policy Areas

The fraction of programs that have reliability growth curves has remained relatively constant over time. Approximately 60 percent of programs with FY 2012 or 2013 approved TEMPs link their reliability growth goal to an OT event.

2.2 *Differences Across Lead Services*

Among programs with FY 2012 or 2013 TEMP approvals, all Services are generally following guidance to:

- Establish a reliability growth or improvement strategy and describe it in the TEMP
- Incorporate reliability growth curves into the TEMP
- Use reliability metrics to ensure growth is on track to achieve requirements.

Army and Navy programs show improvement in implementing the following RAM policies:

- Establishing a reliability growth or improvement strategy (since July 2008, more than 80 percent of Air Force programs have had a reliability growth or improvement strategy)
- Having reliability growth curves and documenting them in the TEMP
- Calculating reliability growth potential.

A larger fraction of Army and Navy programs with FY 2012 or 2013 TEMPs establish and link intermediate goals to the reliability growth curve compared to the Air Force. As shown in Figure 2, Army programs were more likely to link reliability growth goals to OTs compared to the other Services.

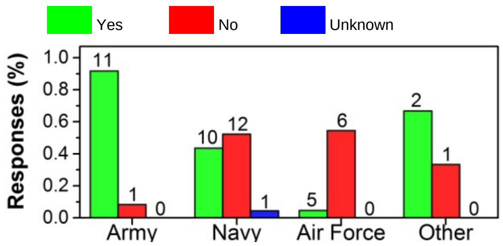

Figure 2 – Fraction Fraction/Number of Responses Indicating Whether the Program Links their Reliability Growth Goal to OT by Lead Service

3 RECOMMENDATIONS

Survey results suggest that military systems should carry out the following activities to improve their chance of meeting reliability requirement in OT:

- Establish OT entrance criteria and ensure these criteria are met prior to proceeding on to the next test phase.
- In accordance with existing USD(AT&L) policy, ensure that that reliability growth curves are stated in a series of intermediate goals and tracked through fully integrated, system-level test and evaluation events until the reliability threshold is achieved.
- Ensure that reliability growth curve assumptions are based on realistic inputs from systems engineering.
- Review the adequacy of requirements to ensure they are achievable.
- Updating reliability growth curves as needed.
- Ensure that enough test time is resourced to support an evaluation of the reliability requirement(s).

REFERENCES

1. Title 10, United States Code, Section 139, “Director Operational Test and Evaluation.”
2. “Director, Operational Test and Evaluation FY 2013 Annual Report,” January 2014.
3. “Defense Acquisition Guidebook,” section 9.5.5, October 2012.
4. DOT&E Memo to the USD(AT&L), “Reliability Survey of Select Acquisition Programs on DOT &E Oversight,” October 30, 2013.

BIOGRAPHIES

Jonathan L. Bell, PhD
Institute for Defense Analyses
Operational Evaluation Division
4850 Mark Center Drive
Alexandria, VA 22310 USA

e-mail: jlbell@ida.org

Jonathan L. Bell is a Research Staff Member at the Institute for Defense Analyses (IDA). He earned his doctoral degree in Materials Science and Engineering at the University of Illinois at Urbana Champaign in 2008 and his bachelor’s degree in Materials Science at Carnegie Mellon University. Dr. Bell’s work at IDA is focused on operational testing of ground vehicle systems with a specific emphasis on reliability.

Matthew R. Avery, PhD
Institute for Defense Analyses
Operational Evaluation Division
4850 Mark Center Drive
Alexandria, VA 22310 USA

e-mail: mavery@ida.org

Matthew Avery, a Research Staff Member at the Institute for Defense Analyses (IDA), earned a Master of Science and doctoral degree in Statistics from the North Carolina State University. Dr. Avery’s work at IDA focuses on statistical aspects of operational test and evaluation.

Michael C. Wells
Institute for Defense Analyses
Operational Evaluation Division
4850 Mark Center Drive
Alexandria, VA 22310 USA

e-mail: mwells@ida.org

Michael C. Wells, a Research Staff Member at the Institute for Defense Analyses (IDA), earned a Master of Science degree in Operations Research from the University of California, Berkeley, a Master of Science in Information Management from Marymount University, and a bachelor’s degree from the United States Military Academy. Mr. Wells joined IDA following 24 years of service with the US Army. Mr. Wells’ work at IDA has focused on operational testing of Army fire support systems.